Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.5, No.14, 2014

Estimation of Generalizability Coefficients: An Application of Structural Equation Modeling

Tuncay Öğretmen[1*] Tülin Acar[2]
1.Education Faculty, Ege University, İzmir, Turkey
2.Parantez Education Research Consultancy Center, Ankara, Turkey
* E-mail of the corresponding author: togretmen@gmail.com

Abstract
The aim of this study is to provide guidance as to the interpretation of results of structural equation model of the variance components according to the generalizability theory. According to the findings, It is observed that G coefficients have been very close to each other which calculated with both LISREL and SPSS programs for both data sets. It has been observed that there is no any significant difference between G coefficients when this difference has been tested with Fisher's Z test. In other words, the estimations of G coefficients that calculated with various programs have been found similar.
Keywords: generalizability coefficiency, reliability, SEM, EduG

1. Introduction

One of the main problems of measuring area in the education and psychology is that measured points represent the actual scores to what extent. In other words, the main issue is to reduce the random errors or to assume the actual scores. The concept of true score in the classic text theory and the concept of universe score in generalizability theory are latent traits in item response theory and true score cannot be observed in all theories, but it is tried to measure with individual's response to the items (Lord,1980; Hambleton ve Swaminathan, 1985;Brennan, 2000). Indeed, it is tried to reduce the measurement errors which are in all measurement theories and to increase the awareness of error sources.

The main premise of classic test theory consists of observed score (X), true score (T) and random errors (E) (Gulliksen, 1950). According to this theory, there is only one error source. However, in fact, the source of the errors that mixed results of measuring has many versions.

X= T + E

X: Observed score

T: True score

E: Random error

Once again, according to classic test theory, reliability index is explained as the square of the correlation between Observed scores and true scores, or the ratio between variance of true scores and variance of Observed scores (Gulliksen, 1950).

$$R^2(X,T) = \frac{\sigma(T)}{\sigma(X)}$$

As the variance of true scores couldn't be known, reliability index has been hypothetic. As the reliability can be estimated by many methods in practice, reliability coefficient has been significant instead of reliability index (Baykul, 2000). According to Crocker and Algina (1986), reliability guess methods are evaluated as "Methods that are multiple and based on a single practice." Difference for reliability guess methods has differentiated the source of measuring errors, and the meaning of the reliability in explaining (internal consistency, consistency, stability).

While it is assumed that measuring errors are at the same levels for all individuals in the classic test theory, measured errors of the individuals have changed in the item reaction theory. However, measuring errors are not the same for all individuals in the congeneric test theory that is a sub-theory of classic test theory (Jöreskog, 1971). Yet, in the generalizability theory based on classic test theory and variance analysis (ANOVA), multiple sources of the measuring errors are evaluated (Brennan, 2005; Webb, Shavelson and Haertel, 2006).

Term of random error takes place in the classic test theory (X= T + E) is modeled by multiple error sources as k pieces error source in the genealizability theory (Brennan, 2011).

$$X = T + E_1 + E_2 + E_3 + ... + E_k$$

X: Observed score

T: True score

E: error source as k pieces (item, pointer, time…etc.)

We can work with small samples both in classic test theory and generalization theory. The most important difference between these two theories is their statistical modeling and approaches to error notion (Suen and Lei, 2007). Generalization theory considers the error sources as systematic and unsystematic. Error sources are called variability source (facet) in the generalization theory, and multiple variability sources (pointers, items, duties, time, test form) are added in statistical model (Brennan, 1992).

There are two studies of generalization theory: G: to be generalized and K: to decide (decision) (Brenann, 1992). In the G study, possible error sources are considered, and several figures are formed depending on several variance analyses in solving (Crocker and Algina, 1986). As for K study, it is focused on a purpose to give a special decision on the data obtained (Brenann, 1992). Effectiveness of the alternative figures are evaluated in order to minimize the errors, and maximum reliability (Webb, Shavelson and Haertel:14). The concept of the reliability coefficient that calculated in the classical test theory must be considered as "genetalizability coefficient" in generalizability theory (Brenann, 1992; Webb ve Haertel,2006). Practical meaning of this generalizability coefficient can be determined in terms of determining of generalizability coefficient more than one for G and K studies which have been established with different research designs.

When the generalization studies in the literature are analyzed, the following studies draw the attention. Atılgan (2005) has benefited from the results of generalization theory in the inter-pointer reliability study in his research. Deliceoğlu and Demirtaşlı (2005) have studied the situations in estimating the reliability of measuring the possible error resources in measuring the football competences depending on generalization theory, and have found that classic test and generalization theory have similar results. Güler (2009) has compared generalizability and reliability coefficients acquired in generalization and decision studies with results of SPSS and GENOVA packages. Yelboğa and Tavşancıl (2010) have analyzed levels of reliability coefficients estimated by classic test theory and G theory, and the relations between them in the work performance-scale used in different times, and have seen that classic test theory and G theory have had similar results. Yılmaz Nalbantoğlu and Gelbal (2011) have analyzed both performance scoring of there pointers together and rotating situation by generalization theory, and then the results have been compared. Güler (2011) has compared reliability on non-random reliable data according to generalization and classic test theory, and has obtained very low values. Anıl and Büyükkıdık (2012) have included results of the Generalization and Decision studies in practicing mixed figure that has considered variability sources as individual, class, pointer, duty and criterion. Performance evaluation has been solved by generalization theory, again, in the study where logistic regression analyzing technique has been suggested to researches in the cases they benefit from generalization theory in order to determine consistency between the pointers sensitively and in detailed, on the other hand they want to have general information without the details about the consistency between the pointers, and they also want to demonstrate existence or absence of the consistency instead of its level, in the measurements consisting of multiple pointers such like performance measurements (Çakıcı Eser and Gelbal, 2012).

The variance components' analysis and reliability estimates of it have been commonly used in the behavioral sciences by means of structural equation models in the international literature (Schoonen,2005; Raykov ve Marcoulides, 2006; Geldhof, Preacher ve Zyphur,2013). The aim of this study is to provide guidance as to the interpretation of results of structural equation model of the variance components according to the generalizability theory. In other words, in estimating the parameters related to generalizability theory, it is the presentation that how estimation of structural equality model is done as an alternative to SPSS and EduG programs At the same time, another purpose of this study is to analyze consistency between G coefficient estimated by structural equality model and SPSS and EduG programs.

2. Method

2.1 Data Collection

Three variability sources such as individual (B), material (M) and Pointers (P) have been discussed in this study. A modeling has been done in nestled pattern within the materials (B*M:P).

SPSS 15.0, EduG 6.0 and LISREL 8.54 programs have been used in the estimation of G coefficients. O'Connor's (2006) script has been used in the estimation of G coefficient in SPSS; and Schoonen'(2005) script has been used in the estimation of G coefficient with Structural Equation Model. Structural Equation Model's script for research data has shown in Appendices 1 and 2.

2.1 Data Analyze

Three variability sources such as individual (B), material (M) and Pointers (P) have been discussed in this study. A modeling has been done in nestled pattern within the materials (B*M:P).

SPSS 15.0, EduG 6.0 and LISREL 8.54 programs have been used in the estimation of G coefficients. O'Connor's (2006) script has been used in the estimation of G coefficient in SPSS; and Schoonen'(2005) script has been used in the estimation of G coefficient with Structural Equation Model. Structural Equation Model's script for research data has shown in Appendices 1 and 2.

3. Findings

Generalizability coefficiency has been calculated as shown below because the aim of this study is to provide guidance as to the interpretation of results of structural equation model of the variance components according to the generalizability theory (Shoonen, 2005).

$$E^2_{\rho\delta} = \frac{\sigma^2_p}{\sigma^2_p + \sigma^2_\delta} \text{ ve } \sigma^2_\delta = \frac{\sigma^2_{pt}}{n_t} + \frac{\sigma^2_{pr}}{n_r} + \frac{\sigma^2_{prt,e}}{n_t n_t} \quad (1)$$

σ^2_δ Error variances in the equation; σ^2_p, σ^2_{pt}, σ^2_{pr} $\sigma^2_{prt,e}$ estimated variances; n_r and n_t are pointer numbers and material numbers. Variance estimations related to the individual and pointers in the Phi Matrix of LISREL's output file have been found. The average of error terms (Theta- Delta) creates the variance of the residuals.

Variance component for the individuals that estimated with structural equation model from first data set consists of 50 students is $\sigma^2_p = 0.33$; variance component for pointers is $\sigma^2_r = (0.55 + 0,34 + 0,52)/3$; variance for residuals is $\sigma^2_{p:pri,e} = (0,50 + 0,10 + 0,31 + 0,67 + 0,43 + 0,65 + 0,44 + 0,35 + 0,30 + 0,30 + 0,66 + 0,25 + 0,47 + 0,71 + 0,66)/15$. 0,639 value has been obtained when G coefficient has been calculated by using (1) equation.

Variance component for the individuals that estimated with structural equation model from second data set is $\sigma^2_p = 0.22$; variance component for pointers is $\sigma^2_r = (0.48 + 0,06 + 0,10)/3$; variance for residuals is $\sigma^2_{p:pri,e} = (-0,13 - 0,14 + 0,33 + 0,34 + 0,56 - 0,02 - 0,02 + 0,07 + 0,07 + 0,23 - 0,08 - 0,08 + 0,15 + 0,14 + 0,15)/15$. 0,738 value has been obtained when G coefficient has been calculated by using (1) equation.

The results related G coefficients that estimated for the both data documents with different software programs are shown in Table 1.

Table 1. Estimations of G Coefficient

	Individual	**Item**	**Pointer**	**YEM**	**SPSS**	**EduG**
1.Data	50	5	3	0,639	0,640	0,740
2.Data	20	5	3	0,738	0,737	0,740

According to the findings, It is observed that G coefficients have been very close to each other which calculated with both LISREL and SPSS programs for both data sets. The calculated G coefficients have been transformed into Fisher's z-statistic and the meaning of difference between G coefficients has been tested with Z test and the results are shown in Table 2.

Table 2. The meaning of difference between G coefficients

		SPSS	**EduG**
1.Data	YEM	-0.008 (p=0,993)	-0,940 (p=0,346)
	SPSS		-0,932 (p=0,351)
2.Data	YEM	0,006 (p=0,994)	-0,013 (p=0,989)
	SPSS		-0,019 (p=0,984)

It has been observed that there is no any significant difference between G coefficients when this difference has been tested with Fisher's Z test (p>0.05). In other words, the estimations of G coefficients that calculated with various programs have been found similar.

5. Conclusion

Mathematically, it is impossible to see that variance is negative and variances in the applications have been estimated negatively. If there is a negative variance in the applications, G coefficients can be estimated with negative variances and structural equation model. But, negative variances in the estimations of G coefficient that made by SPSS and EduG programs have taken as zero. Even though negative variances in YEM analysis have been obtained, the estimation of G coefficient can be evaluated as an advantage.

It is possible to estimate the variance components of G coefficient with the least square methods which are maximum likelihood and weighted with structural equation mode and relevant software gives practitioners the opportunity to use different estimation methods (Shoonen, 2005). On the other hand, G coefficients can be estimated in the programs that analyzed with structural equation model but an output is not produced for decision studies. For this reason, it can be said that SPSS and EduG programs are more useful.

In this study, estimations have been made over data sets where multi-category scoring was made. Therefore, data that made by the binary (dichotomous) scoring will provide important information in the comparison of the results.

Actual application data have been used in the study. Simulation workings that have different sample and pointers have been carried out and the results of estimation of G coefficient can be evaluated by using structural equation model and simulation workings.

References

Anıl, D. ve Büyükkıdık, S.(2012). Genellenebilirlik Kuramında Dört Facetli Karışık Desen Kullanımı İçin Örnek Bir Uygulama. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi. 3(2), 291-296*

Atılgan, H. (2005). A sample Aplication for Generalizability Theory and Interrater Reliability. *Eğitim Bilimleri ve Uygulama*, (4),7.

Ayhan, Y. (2010). İlköğretim 5. sınıf Türkçe ders kitabı resimlemelerinin biçim ve içerik açısından 11 yaş grubuna uygunluğu ve aynı yaş grubu öğrencilerin yaptığı resimlemelerle karşılaştırılması. Yayınlanmamış Yüksek Lisans Tezi, Gazi Üniversitesi, Ankara

Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara: ÖSYM Yayınları

Brennan, R. L. (1992). *Elements of Generalizability Theory*. New York: Springer-Verlog.

Brennan, R. L. (2005). Generalizability Theory. *Educational Measurement: Issues and Practice.* Vol.11 pp. 27-34.

Brennan, R. L. (2011). Generalizability theory and classical test theory. *Applied Measurement in Education, 24*, 1-21.

Crocker, L. ve Algina, J. (1986). Introduction to Classical and Modern Test Theory. Fort Worth: Holt, Rinehart and Winston.

Çakıcı Eser, D. ve Gelbal, S. (2012). Genellenebilirlik Kuramı Ve Lojistik Regresyona Dayalı Hesaplanan Puanlayıcılar Arası Tutarlığın Karşılaştırılması. *Kastamonu Eğitim Dergisi.* 21 (2),423-438.

Çıracıoğlu, Y. (2013). Üç boyutlu görsel sanatlar çalışmalarının, beyin felci (CP) vakasında devinsel becerilerin gelişimine etkisi. Yayınlanmamış Yüksek Lisans Tezi, Gazi Üniversitesi, Ankara

Deliceoğlu, D. ve Çıkrıkçı Demirtaşlı, N. Futbol Yetilerine İlişkin Dereceleme Ölçeğinin Güvenirliğinin Genellenebilirlik Kuramına ve Klasik Test Kuramına Dayalı Olarak Karşılaştırılması, *Spor Bilimleri Dergisi Hacettepe J. of Sport Sciences* 2012, 23 (1), 1–12

Geldhof, G.J., Preacher, J.P.& Zyphur, M.J. (2013). Reliability Estimation in a Multilevel Confirmatory Factor Analysis Framework. *Psychological Methods.* 1-30

Gulliksen, H.O (1950). *Theory of Mental Tests*. New York: John Wiley and Sons, Inc.

Güler, N. (2009). Generalizability Theory and Comparison of the Results of G and D Studies Computed by SPSS and GENOVA Packet Programs. *Education and Science*, 34, 154, 93-103.

Güler, N. (2011). Rasgele Veriler Üzerinde Genellenebilirlik Kuramı ve Klasik Test Kuramına Göre Güvenirliğin Karşılaştırması. *Education and Science*, 36, 162, 225-234.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests.*Psychometrika, 36*, 109–133.

Mushquash, C., & O'Connor, B. P. (2006). SPSS, SAS, and MATLAB programs for generalizability theory analyses. *Behavior Research Methods,* 38(3), 542-547

Raykov, T. & Marcoulides, G. A. (2006). Estimation of Generalizability Coefficients Via a Structural Equation Modeling Approach to Scale Reliability Evaluation. *International Journal of Testing*. 6(1), 81–95

Shoonen, R. (2005). Generalizability of writing scores: an application of structural equation modeling. *Language Testing.*22 (1),1-30.

Suen, H. K. & Lei, P.W. (2007). Classical versus Generalizability Theory of Measurement. *Educational Measurement,* 4. http://suen.ed.psu.edu/~hsuen/pubs/Gtheory.pdf, en son 21 Eylül 2008'de alınmıştır.

Webb, N. M., Shavelson, R. J., & Haertel, E. H. (2006). Reliability Coefficients and Generalizability Theory. *Handbook of Statistics*, Vol. 26 pp. 81-124.

Yelboğa, A**.** ve Tavşancıl, E. (2010). Klasik test ve genellenebilirlik kuramına göre güvenirliğin bir iş performansı ölçeği üzerinde incelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri*. *10*(3). 1825-1854.

Yılmaz Nalbantoğlu, F. ve Gelbal, S. (2011). İletişim becerileri istasyonu örneğinde Genellenebilirlik

Kuramı'yla farklı desenlerin karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 41*, 509-518.

Appendice 1: Estimation coefficient G for the first data file In the SEM

```
da ni=15 no=50 ma = cm
ra
 4 3 3 3 3 5 5 3 3 4 2 4 3 2 2
 5 5 5 5 5 5 5 5 5 5 2 3 2 3 3
 4 4 3 4 4 3 3 3 4 4 2 2 2 1 1
 3 2 3 4 4 4 3 4 4 4 4 2 4 4 4
 1 2 2 2 2 2 2 4 3 3 1 1 2 2 1
 4 5 4 4 4 3 3 3 3 3 4 4 4 5 4
 4 5 4 5 5 4 4 4 4 4 5 4 4 4 4
 5 5 4 5 5 4 4 4 3 3 4 3 4 2 4
 3 4 3 4 2 4 4 4 4 4 4 4 2 4 2
 3 3 2 3 1 2 3 3 3 2 2 3 2 2 2
 2 3 2 1 2 5 5 5 5 4 4 4 4 4 2
 3 2 3 2 4 3 3 3 2 4 2 2 2 2 2
 4 3 4 3 4 5 4 5 5 5 4 2 4 4 2
 5 3 5 5 5 5 3 5 5 3 4 2 4 4 4
 5 5 5 5 5 5 5 5 5 5 4 5 3 4 3
 5 4 5 5 5 4 5 5 5 5 5 4 5 5 5
 5 4 5 5 5 5 5 5 5 5 5 4 5 5 5
 5 4 5 5 5 5 5 5 5 5 5 4 5 5 5
 5 4 5 5 5 5 5 5 4 4 5 4 5 5 5
 5 4 5 5 5 5 5 5 5 5 5 4 5 5 5
 4 2 2 3 3 5 5 5 5 5 4 5 5 5 5
 4 3 4 4 4 4 2 2 3 3 2 4 4 4 2
 5 4 4 3 4 4 3 4 4 4 2 2 4 4 4
 5 4 3 3 3 5 4 4 4 3 4 4 2 2 2
 4 3 2 2 2 4 3 2 4 5 4 2 2 4 4
 5 3 3 3 3 2 3 2 2 3 2 2 2 2 4
 1 1 2 1 3 4 2 2 2 2 4 2 2 2 2
 4 3 4 3 4 2 2 4 2 2 2 2 4 2 2
 4 4 3 3 4 4 4 3 2 4 4 4 2 2 4
 3 3 5 4 5 4 4 2 4 4 4 4 2 2 4
 5 5 4 5 5 2 2 4 4 4 2 2 4 4 4
 3 4 3 3 3 4 4 4 3 4 5 5 5 5 4
 4 5 5 5 5 5 5 5 5 4 5 4 4 5 5
 2 3 2 2 2 5 4 5 4 5 5 4 4 4 2
 3 4 5 5 4 5 5 4 4 4 2 3 2 2 2
 4 3 3 1 3 3 3 3 3 2 3 4 5 5 4
 5 5 5 5 5 3 4 5 5 5 4 3 4 2 3
 4 3 4 4 4 2 2 2 2 2 5 5 5 5 5
 1 2 2 2 2 5 5 4 4 4 4 3 4 4 4
 3 2 2 3 2 4 3 3 4 3 2 2 2 2 2
 2 3 4 5 3 4 3 4 4 4 3 3 2 3 2
 5 4 4 4 4 3 4 3 3 4 4 3 4 3 3
 4 4 3 5 3 1 2 3 4 2 5 4 4 4 4
 3 4 4 4 5 4 4 4 4 2 4 5 4 5 3
 3 4 4 4 4 3 3 3 5 3 2 3 3 3 3
 4 5 5 5 5 2 3 3 3 4 2 3 3 3 3
 5 4 5 5 5 2 3 3 3 3 3 5 5 5 5
 5 4 4 4 4 4 5 4 4 4 5 4 5 5 5
 5 4 5 5 5 5 5 5 5 5 5 4 4 4 4
 5 4 4 5 5 5 4 4 4 3 5 4 4 5 5
mo nx = 15 nk = 4 ph = di
lk
person r1 r2 r3
va 1 lx (1,1) lx (2,1) lx (3,1) lx (4,1) lx (5,1) lx (6,1)
```

Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.5, No.14, 2014

```
va 1 lx (7,1) lx (8,1) lx (9,1) lx (10,1) lx (11,1) lx (12,1) lx (13,1) lx (14,1) lx (15,1)
va 1 lx (1,2) lx (2,2) lx (3,2) lx (4,2) lx (5,2)
va 1 lx (6,3) lx (7,3) lx (8,3) lx (9,3) lx (10,3)
va 1 lx (11,4) lx (12,4) lx (13,4) lx (14,4) lx (15,4)
pd
ou me = uls se
```

Appendice 2: Estimation coefficient G for the second data file In the SEM

```
da ni=15 no=20 ma = cm
ra
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
3 3 2 2 2 2 2 2 2 2 2 2 2 2 2
3 3 2 2 2 2 2 3 3 3 2 2 2 2 2
2 2 2 3 3 2 2 2 2 2 2 2 2 2 2
2 2 2 3 3 2 2 2 2 3 2 2 2 2 2
4 4 5 5 5 3 3 3 3 3 2 2 2 2 2
4 4 5 5 5 3 3 3 3 3 3 3 3 3 3
3 3 3 3 4 3 3 3 3 3 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
4 4 4 4 5 3 3 3 3 4 2 2 2 2 3
2 2 3 3 3 3 3 3 3 4 3 3 3 3 3
3 3 3 3 3 3 3 3 3 4 3 3 3 3 3
2 3 3 4 4 3 3 4 4 4 3 3 4 4 4
2 2 2 2 2 3 3 3 3 3 3 3 4 4 4
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
2 2 4 4 4 3 3 3 3 3 3 3 3 2 2
2 2 2 2 2 2 2 2 2 3 2 2 2 2 2
2 2 2 2 2 2 2 2 2 3 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2

mo nx = 15 nk = 4 ph = di
lk
persoon r1 r2 r3
va 1 lx (1,1) lx (2,1) lx (3,1) lx (4,1) lx (5,1) lx (6,1)
va 1 lx (7,1) lx (8,1) lx (9,1) lx (10,1) lx (11,1) lx (12,1) lx (13,1) lx (14,1) lx (15,1)
va 1 lx (1,2) lx (2,2) lx (3,2) lx (4,2) lx (5,2)
va 1 lx (6,3) lx (7,3) lx (8,3) lx (9,3) lx (10,3)
va 1 lx (11,4) lx (12,4) lx (13,4) lx (14,4) lx (15,4)
pd
ou me = uls se
```